Research Article

The Importance of the Model Choice for Experimental Semivariogram Modeling and Its Consequence in Evaluation Process

Alessandro Mazzella\(^1\) and Antonio Mazzella\(^2\)

\(^1\) Department of Social Science, University of Cagliari, 09123 Sardinia, Italy
\(^2\) Department of Civil and Environmental Engineering and Architecture, University of Cagliari, 09123 Sardinia, Italy

Correspondence should be addressed to Alessandro Mazzella; amazzella@unica.it

Received 8 January 2013; Revised 14 June 2013; Accepted 24 July 2013

Academic Editor: Sergio Nardini

Copyright © 2013 A. Mazzella and A. Mazzella. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Geostatistics was created during the second half of 20th century by Georges Matheron, on the basis of Danie Krige’s and Herbert Sichel’s theories [1–5]. The purpose of this new science was to achieve an optimal evaluation of mining ore bodies. The interest in geostatistical tools has grown, and nowadays its techniques are applied in many branches of engineering where data analysis, interpolation, and evaluation are necessary. This paper presents an overview of the geostatistics approach in data analysis and describes each operative step from experimental semivariogram calculation to kriging interpolation, focusing and underlining the experimental semivariogram modeling step. To help any data analysts during geostatistical analysis process, an innovative geostatistical software was created. This new software, named “Kriging Assistant” (KA) and developed within the Department of Geoengineering and Environmental Technologies University of Cagliari, is able, with a marginal support of the user, to produce 2D and 3D grids and contour maps of sampled data. A comparison between kriging results obtained by KA and two of the most common data analysis softwares (Golden Software Surfer and ESRI Geostatistical Analyst for ArcMap) is presented in this paper. Reported data showed that KA minimizes interpolation errors and, for this reason, provides better interpolation results.

1. Introduction

Geostatistics was born during the last century in the mining field. Georges Matheron, on the basis of Danie Krige’s and Herbert Sichel’s theories [1–5], created new tools for the evaluation of mineral deposits; Bertil and Gandin provided the same tools in meteorological and forestal fields [6, 7]. This new approach was based on the “regionalized variables” theory [8]: a new type of variable influenced by its position within a mineralized “region.” According to this theory, a “regionalized variable,” schematically represented in Figure 1, could be defined by

\[ z(x) = m(x) + k(x), \]

where \( m(x) \) is the regionalized component and \( k(x) \) represents the random component that explains the local effects.

Since then a lot of progress has been made in the development of geostatistics techniques of data analysis and interpolation. For this reason, geostatistics has become an extremely powerful tool for studying and evaluating space- and/or time-related phenomena, and in the present days, its own techniques are implemented in all of the most popular data analysis softwares.

Presently, geostatistics supplies a collection of powerful techniques that address the study of spatial correlation between experimental values of a specific variable in order to estimate values in unknown points inside the phenomenon existing domain.

Geostatistical analytic approach is mainly based on two different operative steps:

(i) semivariogram calculus and modelling;

(ii) kriging interpolation technique.
2. Semivariogram Calculus and Modeling

Semivariogram, whose equation is shown in (2), is the geostatistical tool for studying the relationship between collected data in function of distance and direction,

\[ \gamma(h) = \frac{1}{2} n(h) \frac{\sum (z(x+h) - z(x))^2}{n(h)}. \]  

Semivariogram interpretation, whose schematic representation is shown in Figure 2, mainly concerns the definition of the variability model that best fits the experimental curve shape. Subsequently, it becomes possible to proceed with the variographic parameters evaluation process. These operative steps are necessary to uniquely identify the phenomenon law with a theoretical curve. Referring to Figure 2, variographic parameters are

(i) the “nugget effect” \( \gamma_0 \), which represents the value of the initial variability;

(ii) the “range” \( r \), which represents the distance (measured in horizontal axis units) beyond which the data becomes totally independent;

(iii) the “sill” \( c \), which represents the maximum variability value reached at a distance equal to the range; and

(iv) the “model equation” or the theoretical curve which best fits the experimental curve shape (Figure 3).

Referring to Figure 3, most common theoretical models are

(i) “spherical model,” also known as “Matheron model” is represented by a simple polynomial expression: its trend shows a steady increase up to a distance equal to the “range” and then reaches a plateau;”

\[ \gamma(h) = \begin{cases} 
\gamma_0 & r = |h| \leq a \\
\gamma_0 + c & r = |h| > a 
\end{cases} \]  

(ii) “exponential model” (4): because this model reaches the threshold value with an asymptotic trend, its range is defined as the distance at which the \( \gamma \)-value corresponds to 95% of the threshold value; consider

\[ \gamma(h) = \begin{cases} 
\gamma_0 (1 - e^{-|r|/a}) & r = |h| \leq a \\
\gamma_0 & r = |h| > a 
\end{cases} \]  

(iii) “gaussian model” (5): this model is used in case of extremely continuous phenomena; consider

\[ \gamma(h) = \begin{cases} 
\gamma_0 (1 - e^{-2r^2/a^2}) & r = |h| \leq 0 \\
\gamma_0 & r = |h| > a 
\end{cases} \]  

(iv) “linear model” (6): this model is the simplest one without a sill and is characterized by a linearly increasing trend with increasing distance; consider

\[ \gamma(h) = \gamma_0 |h|, \quad r = |h| \leq 0; \]  

\[ \gamma(h) = c, \quad r = |h| > 0. \]
(v) “power model” (7): this model is characterized by parabolic trend,
\[ \gamma(h) = cr^\omega; \quad r = 0 < \omega < 2. \]  

Theoretical model equation and variographic parameters values will be used by kriging algorithm to calculate the unknown values.

### 2.1. Data Influence on Experimental Semivariogram Shape

An optimal choice of the semivariogram model is an important point for a good data evaluation process. Since semivariogram expresses the relationship between measured values themselves, it is obvious that the model recognition strongly influences all the evaluation process.

Some phenomena show anisotropic behaviour (i.e., the behaviour of the variables that characterize them is not uniform in all directions). Experimental semivariogram calculus along different directions can highlight the presence of anisotropy in the spatial distribution of the “regionalized variable” as shown in Figures 4 and 5.

Similar considerations can be made about the influence of sampled data on the “sill” and the “nugget effect.”

In Figure 6, a particular kind of regular grid sampling is shown. Referring to Figure 6, on the basis of their position, couples of data points can be classified in the following types:

(i) both points are located outside the “regionalized” area (type “I” couple);
(ii) only one point is located outside the “regionalized” area (type “II” couple);
(iii) both points are located within the “regionalized” area (type “III” couple).

These three types of couples contribute and influence in a different way the experimental semivariogram shape and, consequently, the variographic parameter values.

In particular type “I” and type “II” couples influence the “nugget effect” value and the “sill” value. This is because these types of couples act as independent data, and thus they contribute to the uncertainty of the correlation law which is represented by the “nugget effect” and the “sill” values.

On the contrary, type “III” couples contribute to the definition of the semivariogram shape within the range being tied by the law of the phenomenon of which the semivariogram is a graphical representation.

For these reasons the semivariogram structure shows

(i) an increasing trend with the increasing distance between the points of the couple;
(ii) a value near the origin which is a function of data inhomogeneity;
(iii) a final value as high as the maximum difference between data.
3. Kriging Interpolation Technique

Kriging interpolation considers only the sampled data which exercise a real influence on the unknown point, as shown in Figure 7, through the parameters coming from the model semivariogram.

Kriging interpolation technique calculates unknown values using

\[ z_x = \sum_{i=1}^{n} \lambda_i z_i, \]  

(8)

where \( z_i \) are known because they are the sampled values within the range of influence while the weighting values \( \lambda \) are calculated as the solutions of the kriging equation system shown in (9), where \( y(z_i, z_j) \) are calculated by the equation of the chosen semivariogram model.

From the relationship between model equation and lambda, the strong importance of the variographic modelling step on the kriging results is evident. In fact the absence of a standard technique for variogram modelling implies that kriging results might be influenced by personal interpretations of the data analyst.

Moreover, it is important to emphasize that the evaluation process does not end with kriging itself, but, at the end of interpolation step, it is important to proceed with results validation:

\[
\begin{align*}
\lambda_1 y(z_1, z_1) + \lambda_2 y(z_1, z_2) + \cdots + \lambda_n y(z_1, z_n) + \mu &= y(z_1, P) \\
\lambda_1 y(z_2, z_1) + \lambda_2 y(z_2, z_2) + \cdots + \lambda_n y(z_2, z_n) + \mu &= y(z_2, P) \\
\vdots \\
\lambda_1 y(z_n, z_1) + \lambda_2 y(z_n, z_2) + \cdots + \lambda_n y(z_n, z_n) + \mu &= y(z_n, P) \\
\lambda_1 + \lambda_2 + \cdots + \lambda_n &= 1.
\end{align*}
\]

(9)

From the relationship between model equation and \( \lambda_i \), the strong importance of the variographic modelling step on the kriging results is evident. In fact the absence of a standard technique for variogram modelling implies that kriging results might be influenced by personal interpretations of the data analyst.

4. Results Validation: “Krigie Diagram”

At the end of an estimation process the "krige diagram" is useful to appreciate the goodness of the estimation process.

In Figure 8, where the "krige diagram" is shown, the validation of an estimation process is shown. In this representation, each real data, known only "a posteriori," is plotted versus its interpolated value. It is evident that in the case of ideal interpolation, all data would lie along a 45° line expressing a perfect correspondence between true and interpolated values. In real cases, all the data spread within an ellipse.

Referring to Figures 8(a) and 8(b), once defined a cut-off value (e.g. usually an economic limit value), krigie diagram is divided into 4 quadrants:

(i) S-W (named “quadrant 1”), where both calculated and measured values are lower than cut-off (correct estimation process);

(ii) N-W (named “quadrant 2”), where calculated values are lower than cut-off values but measured ones are higher (bad but not danger estimate);

(iii) N-E (named “quadrant 3”) where both calculated and measured values are higher than cut-off (correct estimation process);
(iv) S-E (named “quadrant 4”) where estimated values are higher than cut-off but real values are lower (bad and very danger estimate).

During an evaluation study only estimated data are available; for this reason the areas selected to make production are quadrants 3 and 4. While quadrant 3 is correctly cultivated quadrant 4 is composed by samples with values lower than cut-off with obvious consequence on estimation project.

Consider that during this synthetic analysis we have not considered the estimation technique. In conclusion, to reduce any problem we need to reduce all possible errors; for this reason, we must consider the process of semivariogram modeling [10].

5. Overestimation and Solutions

In practice, the necessity to limit overestimation errors requires a correct execution of the entire evaluation process, that is:

(i) area delimitation and in situ characterization;
(ii) sampling;
(iii) sample chemical analysis;
(iv) statistical results analysis;
(v) geostatistical results analysis (especially: variographic analysis and modelling);
(vi) kriging;
(vii) kriging results validation.

Among the different possible strategies to achieve evaluation errors reduction, our research team has been working since 1990 on developing geostatistical innovative techniques. During the last years, we have been focusing our attention on building novel techniques that are able to assist the evaluator in

(i) the correct choice of the theoretical model for the experimental semivariogram modelling;
(ii) the correct evaluation of the variographic parameters.

5.1. Experimental Semivariogram Modeling. Previously, it has been already said that experimental semivariogram modelling is one of the geostatistics main problems that remain to be solved. In this section some theoretical foundations of procedures implemented into the proposed geostatistical analysis tool will be introduced. This procedure is able to choose...
the variogram model by itself and to calculate variographic parameters: its application can avoid the influence of an incorrect interpretation and limit interpolation errors.

Novel technique developed and already validated for the semivariogram modelling and variographic parameters estimation are based on the construction of a “filtered” experimental semivariogram. This kind of semivariogram is calculated through a process which limits the influence of the low number of couples as shown in the examples of Figure 9 and Table 1.

5.2. Variographic Parameters Evaluation. From the filtered experimental semivariogram [10], it is possible to calculate “nugget effect,” “sill,” and “range” directly from experimental data.

5.2.1. Nugget Effect Calculus. “Nugget effect” calculus is based on a geometrical extrapolation procedure from the first two points of the filtered semivariogram extending the line formed by these points until the intersection with the y-axis, Figure 10.

5.2.2. Range and Sill Calculus. On the basis of the filtered semivariogram it is possible to define the starting point from which the experimental semivariogram curve becomes horizontal, Figure 11. In this way, it is possible to define the range and the "sill.”

6. Kriging Assistant Geostatistical Tool

Kriging assistant (KA) is based upon the experience of: FGAM (Figure 12), a tool for calculating the experimental semivariogram of a dataset, and JCBLOK, a tool for interpolation through Kriging [9].

KA (whose main window is shown in Figure 13) represents an evolution of Carr and Melas’ work. It has been written in Microsoft Visual Basic 6, using the GeoLIBs libraries [13], and it gathers the FGAM and JCBLOK characteristics, but it has been rewritten and implemented in order to be more powerful, user-friendly, and efficient.

Once the user starts KA (by clicking twice on its desktop icon), the software alert window is shown. The message informs the user about the default condition for semivariogram calculus and the instructions for customizing them (Figure 13). Regarding the correct datafile management and the accurate semivariogram calculus, the most important parameters are “dimensionality” and “1, 2, or 3D” (Figure 14).

Through the customization of these two parameters KA will know the sampling mesh features (sampling made along a line, along a gallery, along a bore hole, etc.) and will calculate a mono-, bi-, or three-dimensional semivariogram. In particular, the first parameter regards the datafile reading procedure whereas the second one deals with the calculation algorithm. After these parameters customization KA is ready to import the data and to proceed to the geostatistical estimation.

The current KA version (v3.0/2011) is able to read only two different datafile standards: GeoEAS datafile, csv datafile.

The first one is a dataset introduced since 1991 by the program GeoEAS (geostatistical environmental assessment software) which represents a de facto standard nowadays [14]. This standard features are (Figure 15) the first line which is reserved for the title of the project; the second row which indicates the number of variables stored in the datafile; many rows as number of variables indicating the name and format (according to the FORTRAN syntax) of the data and the sequence of data separated by white space. The other supported standard is the classic “comma delimited.”

Once the datafile has been imported within KA, the sample map is automatically generated (Figure 16). The average time required for the map generation is approximately 0.2 seconds every 4000 records.

The KA sample map represents the position (indicated by red circles) and the measured values (on the right of each red circle) of each sample. Naturally, in the case of a large amount of samples, KA adjusts the map scale to prevent the display of overlapping of values. This simple representation allows the user to perform a quick visual inspection of the mesh sampling integrity. Then the experimental semivariogram calculus is shown, Figure 17, with its semivariogram curve, modeling, and interpretation. This step is primary for the calculus of the parameters needed for the next phase of data kriging.

Once the user selects the technique for the semivariogram modelling [12], KA shows the graphical representation of the theoretical models (in the graphical output window) and the calculated values of “Nugget effect,” “range,” and “sill” (in the kriging configuration panel). At this point, the
7. Kriging Assistant Validation

In order to test KA reliability, a comparison between a geostatistical analysis made by KA and two of the most famous commercial data analysis softwares was made (i.e., Golden Software Surfer and Esri ArcMap Geostatistical Analyst).

A regular 3D-grid equally spaced by 10 meters has been created; it has been overlapped to a 10 meters-Digital Elevation Model of a 500 m × 500 m region (Figure 20). 25 random points were removed and calculated as unknown values by a geostatistical approach with KA, Surfer, and ArcMap Geostatistical Analyst.

In Table 2, a comparison between true values, KA kriged values, Surfer kriged values and ArcMap Geostatistical Analyst kriged values is shown.

8. Conclusion

The primary target of this work was to explain the main advantages of geostatistics, to underline the errors arising from the geostatistical evaluation approach, and to propose
Table 2: Comparison between kriging results and true values.

<table>
<thead>
<tr>
<th>Point_id</th>
<th>Point, z</th>
<th>Surfer interpolation</th>
<th>Geostatistical analyst interpolation</th>
<th>KA interpolation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Point, z</td>
<td>Delta</td>
<td>Delta^2</td>
</tr>
<tr>
<td>162</td>
<td>271.43</td>
<td>271.82</td>
<td>0.39</td>
<td>0.15</td>
</tr>
<tr>
<td>193</td>
<td>146.55</td>
<td>147.29</td>
<td>0.74</td>
<td>0.54</td>
</tr>
<tr>
<td>329</td>
<td>214.69</td>
<td>214.77</td>
<td>0.08</td>
<td>0.01</td>
</tr>
<tr>
<td>411</td>
<td>188.64</td>
<td>188.24</td>
<td>-0.40</td>
<td>0.16</td>
</tr>
<tr>
<td>616</td>
<td>181.56</td>
<td>181.63</td>
<td>0.07</td>
<td>0.00</td>
</tr>
<tr>
<td>645</td>
<td>207.62</td>
<td>207.59</td>
<td>-0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>679</td>
<td>287.69</td>
<td>287.84</td>
<td>0.14</td>
<td>0.02</td>
</tr>
<tr>
<td>877</td>
<td>184.05</td>
<td>183.76</td>
<td>-0.28</td>
<td>0.08</td>
</tr>
<tr>
<td>917</td>
<td>155.31</td>
<td>155.06</td>
<td>-0.25</td>
<td>0.06</td>
</tr>
<tr>
<td>1272</td>
<td>183.97</td>
<td>183.43</td>
<td>-0.54</td>
<td>0.29</td>
</tr>
<tr>
<td>1399</td>
<td>243.56</td>
<td>243.39</td>
<td>-0.17</td>
<td>0.03</td>
</tr>
<tr>
<td>1398</td>
<td>172.85</td>
<td>172.56</td>
<td>-0.29</td>
<td>0.09</td>
</tr>
<tr>
<td>1422</td>
<td>210.67</td>
<td>211.08</td>
<td>0.41</td>
<td>0.16</td>
</tr>
<tr>
<td>1722</td>
<td>278.81</td>
<td>278.57</td>
<td>-0.24</td>
<td>0.06</td>
</tr>
<tr>
<td>1733</td>
<td>236.76</td>
<td>236.81</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>1844</td>
<td>206.62</td>
<td>206.43</td>
<td>-0.18</td>
<td>0.03</td>
</tr>
<tr>
<td>1935</td>
<td>274.15</td>
<td>273.53</td>
<td>-0.62</td>
<td>0.39</td>
</tr>
<tr>
<td>1960</td>
<td>201.38</td>
<td>201.10</td>
<td>-0.28</td>
<td>0.08</td>
</tr>
<tr>
<td>2075</td>
<td>177.02</td>
<td>176.67</td>
<td>-0.35</td>
<td>0.12</td>
</tr>
<tr>
<td>2169</td>
<td>211.48</td>
<td>210.92</td>
<td>-0.56</td>
<td>0.31</td>
</tr>
<tr>
<td>2261</td>
<td>223.21</td>
<td>223.28</td>
<td>0.07</td>
<td>0.00</td>
</tr>
<tr>
<td>2428</td>
<td>204.57</td>
<td>205.09</td>
<td>0.52</td>
<td>0.27</td>
</tr>
<tr>
<td>2499</td>
<td>236.47</td>
<td>235.82</td>
<td>-0.65</td>
<td>0.43</td>
</tr>
<tr>
<td>2508</td>
<td>258.17</td>
<td>257.89</td>
<td>-0.28</td>
<td>0.08</td>
</tr>
<tr>
<td>2598</td>
<td>183.74</td>
<td>183.26</td>
<td>-0.48</td>
<td>0.23</td>
</tr>
<tr>
<td>SUM</td>
<td></td>
<td></td>
<td>-3.15</td>
<td>3.60</td>
</tr>
</tbody>
</table>

Figure 17: On the left, the experimental semivariogram calculus result is shown; on the right, the panel for the automatic interpretation technique selection is shown.

Figure 18: The KA kriging configuration panel with the numerical results of the interpretation and, in blue, the kriging starting button.

Figure 19: KA evaluation graphical results: on the left the “block map” and on the right the “contour map.”

a possible solution. In this setting, a novel software that is able to execute, with a fully automated approach and using techniques not found in any other analytical software, the geostatistical evaluation of “regionalised variables” has been introduced.

The software, called “kriging assistant” (KA), was born on the experience of two American researchers, James R. Mela and Kenneth Carr, who developed two different applications.
for the analysis of geostatistical data (i.e., the “FGAM” to calculate the experimental semivariogram and “Jcblock” for the block kriging procedure application).

However, KA has been developed in a different way than that of Carr and Mela: many parts have been redesigned, rewritten, and adapted in order to produce a stable and user-friendly analytical tool, particularly suitable for the treatment of all types of data. In particular, KA is based on novel tools for automatic modelling of the experimental semivariogram and for high quality graphics realization.

Over the years, technological improvements led to significant increases in computing power (in terms of quantity of the data managed) and speed (in terms of computation time). These aspects, however, did not imply a consequent improvement of quality of the estimates produced. In fact it should be noted that the evaluation process is just the latest step in a sophisticated set of operations, such as:

(i) area delimitation and in situ characterization;
(ii) sampling;
(iii) sample chemical analysis;
(iv) statistical results analysis;
(v) geostatistical results analysis (especially: variographic analysis and modelling);
(vi) kriging;
(vii) kriging results validation.

Each of these operations is inevitably a source of errors that can influence the entire evaluation process results: an erroneous sampling, an insufficient number of data, a reduction of the samples performed incorrectly, laboratory analytical errors, and so on, which will certainly affect the evaluation quality. For this reason, KA, like all other softwares for data analysis, is unable to ensure accurate results if the analyst did not proceed with the utmost care in all previous operation steps.

Another aspect to be considered is that, despite its user-friendly interfaces, KA should be used by a skilled user: in fact in many operations that the program automatically executes (such as the experimental semivariogram modelling or the data kriging) the user must customize partially or entirely the parameters processed by the software.

KA future developments are in progress especially along two main directions:

(i) on one hand, we are working in order to improve the capacity to import data and to export results in different formats (e.g., ∗.xls, ∗.mdb, ∗.jpg, ∗.tiff);
(ii) on the other hand, we are working in order to develop techniques to produce results that are directly manageable within a GIS environment.

The development of special tools for automatic conversion of data type numeric descriptive data would make KA more desirable to a wide range of experts (e.g., engineers, geologist, etc.) working on environmental processes.

Acknowledgment

The authors wish to thank Mrs. Gesualda Galdini for her precious support during all the research work.

References


